

Reconocimiento de Patrones

Version 2022-2

Formulación Metodológica de un Problema de Reconocimiento de Patrones

[Capítulo 1]

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electronica

Universidad Popular del Cesar

Formulación Metodológica

1. Obtención de la Información Original

2. Pre-procesamiento

3. Extracción de Características

4. Normalización de Características

5. Análisis de Características

6. Selección de Características

7. Diseño del Clasificador

8. Evaluación del Desempeño

Durante la explicación seguiremos este ejemplo ...

Mandarinas



Naranjas



¿cómo separar las mandarinas de las naranjas?

Formulación Metodológica

1. Obtención de la Información Original

2. Pre-procesamiento

3. Extracción de Características

4. Normalización de Características

5. Análisis de Características

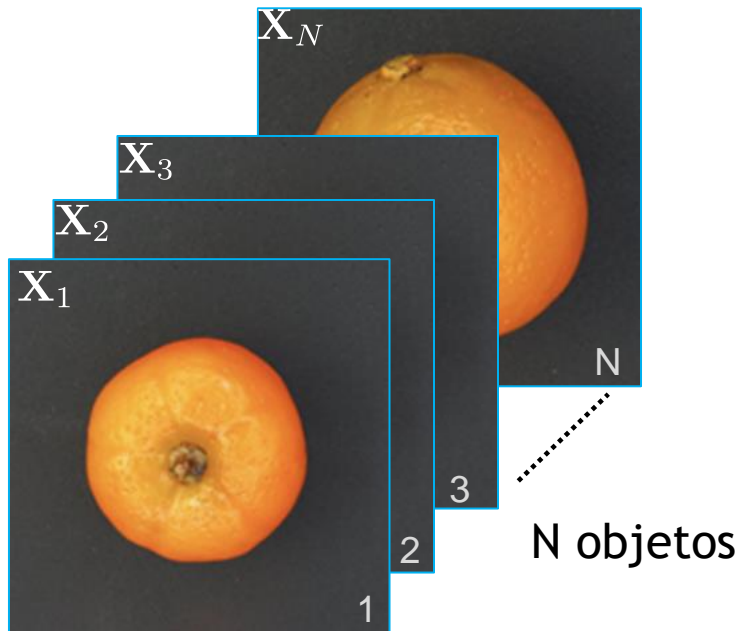
6. Selección de Características

7. Diseño del Clasificador

8. Evaluación del Desempeño

1. Obtención de la Información Original

Los datos (imágenes, señales, etc.) se capturan...



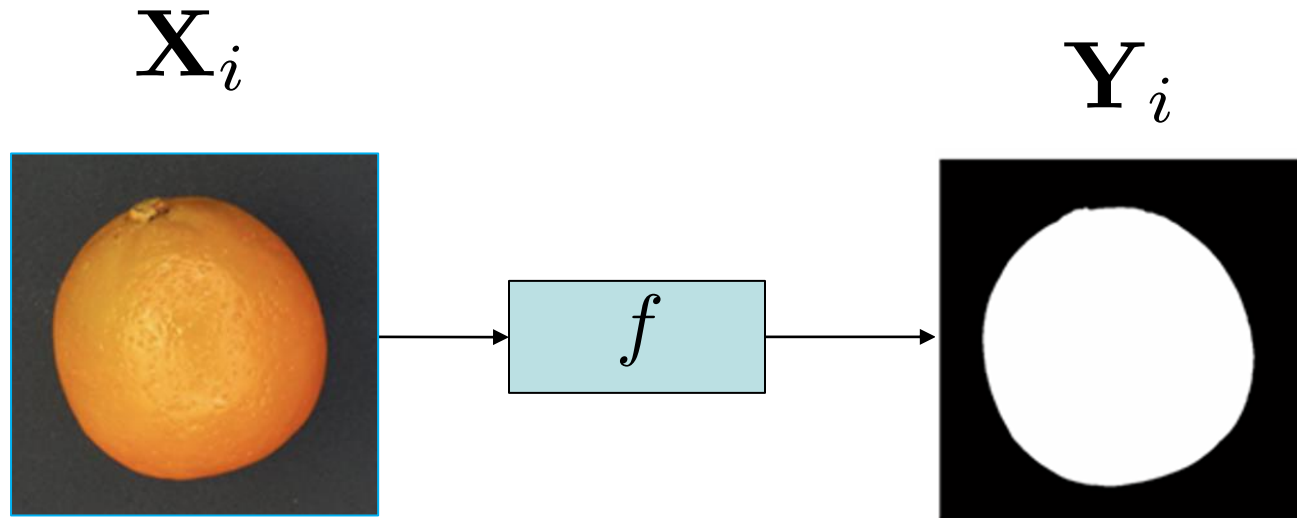
$$\mathbf{X}_i \quad \text{para} \quad i = 1 \dots N$$

Matriz (imagen) del objeto i

Ejemplo: Mandarinas y Naranjas

2. Pre-procesamiento

Los datos se limpian, filtran, segmentan, etc.



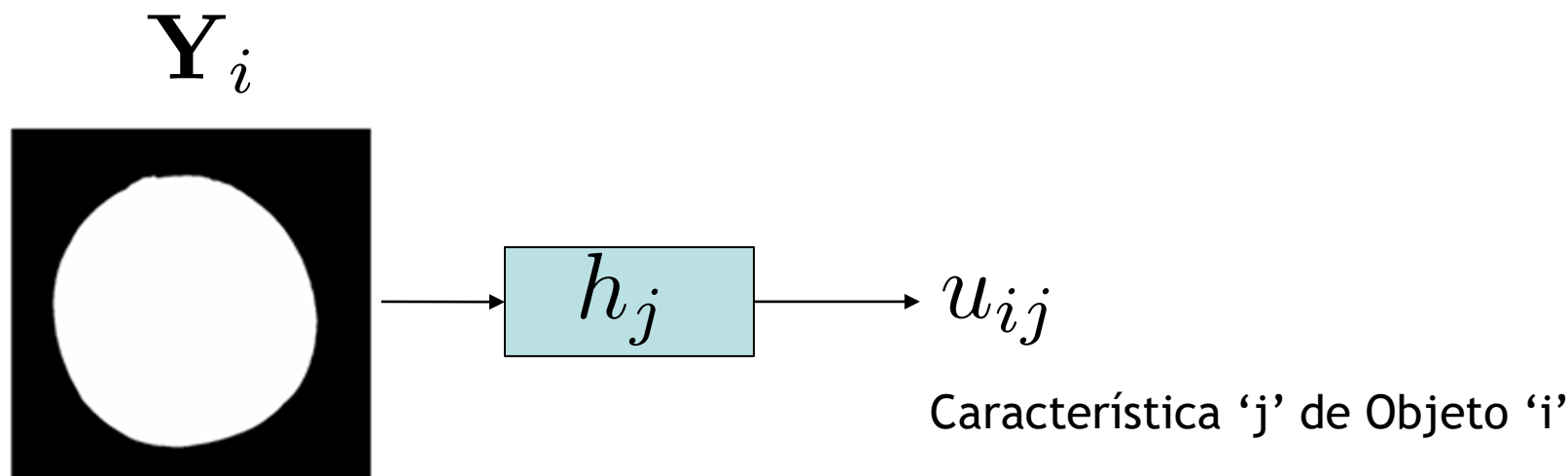
$$\mathbf{Y}_i = f(\mathbf{X}_i) \quad \text{para} \quad i = 1 \cdots N$$

Formulación Metodológica

1. Obtención de la Información Original
2. Pre-procesamiento
- 3. Extracción de Características**
4. Normalización de Características
5. Análisis de Características
6. Selección de Características
7. Diseño del Clasificador
8. Evaluación del Desempeño

3. Extracción de Características

Para cada objeto se extraen M características...



$$u_{ij} = h_j(\mathbf{Y}_i) \quad \text{para} \quad j = 1 \cdots M$$

3. Extracción de Características

Ejemplo: Para las Mandarinas (50) y Naranjas (75) se extrae:

- 1) Área (A)
- 2) Rojo (R)
- 3) Verde (G)
- 4) Azul (B)

En este caso, $N = 50 + 75 = 125$:

Matriz \mathbf{U} de 125×4

| | | \mathbf{u}_1 | \mathbf{u}_2 | \mathbf{u}_3 | \mathbf{u}_4 |
|------------|-----|----------------|----------------|----------------|----------------|
| | | A | R | G | B |
| Mandarinas | 1 | 13234 | 0.27 | 0.40 | 0.11 |
| | : | 12129 | 0.29 | 0.39 | 0.09 |
| | 50 | 11957 | 0.31 | 0.42 | 0.12 |
| Naranjas | 51 | 12911 | 0.35 | 0.38 | 0.13 |
| | : | | | | |
| | 125 | 17288 | 0.30 | 0.73 | 0.10 |

$= \mathbf{U}$

$u_{1,2}$

Formulación Metodológica

1. Obtención de la Información Original
2. Pre-procesamiento
3. Extracción de Características
- 4. Normalización de Características**
5. Análisis de Características
6. Selección de Características
7. Diseño del Clasificador
8. Evaluación del Desempeño

4. Normalización de Características

| | | u_1 | u_2 | u_3 | u_4 | |
|------------|---------|-------|-------|-------|-------|-------|
| | Muestra | A | R | G | B | |
| Mandarinas | 1 | 13234 | 0.27 | 0.40 | 0.11 | = U |
| | : | 12129 | 0.29 | 0.39 | 0.09 | |
| | 50 | 11957 | 0.31 | 0.42 | 0.12 | |
| Naranjas | 51 | 12911 | 0.35 | 0.38 | 0.13 | |
| | : | | | | | |
| | 125 | 17288 | 0.30 | 0.73 | 0.10 | |



Escalamiento: cada columna se normaliza...

$$\mathbf{v}_j = f_n(\mathbf{u}_j) = a_j \mathbf{u}_j + b_j$$

columna normalizada

columna j de U

Normalización:

- MinMax:
min = 0, max = 1
- Mean0:
mean = 0, var = 1

4. Normalización de Características

| | | Muestra | A |
|------------|---|---------|-------|
| Mandarinas | { | 1 | 13234 |
| | | : | 12129 |
| | | 50 | 11957 |
| Naranjas | { | 51 | 12911 |
| | | : | |
| | | 125 | 17288 |

Columna Original

u_1

4. Normalización de Características

| | | Muestra | A |
|------------|---|---------|--------|
| Mandarinas | { | 1 | 0.4981 |
| | | : | 0.3681 |
| | | 50 | 0.3479 |
| Naranjas | { | 51 | 0.4601 |
| | | : | |
| | | 125 | 0.9751 |

Columna Normalizada con MinMax

V_1

4. Normalización de Características

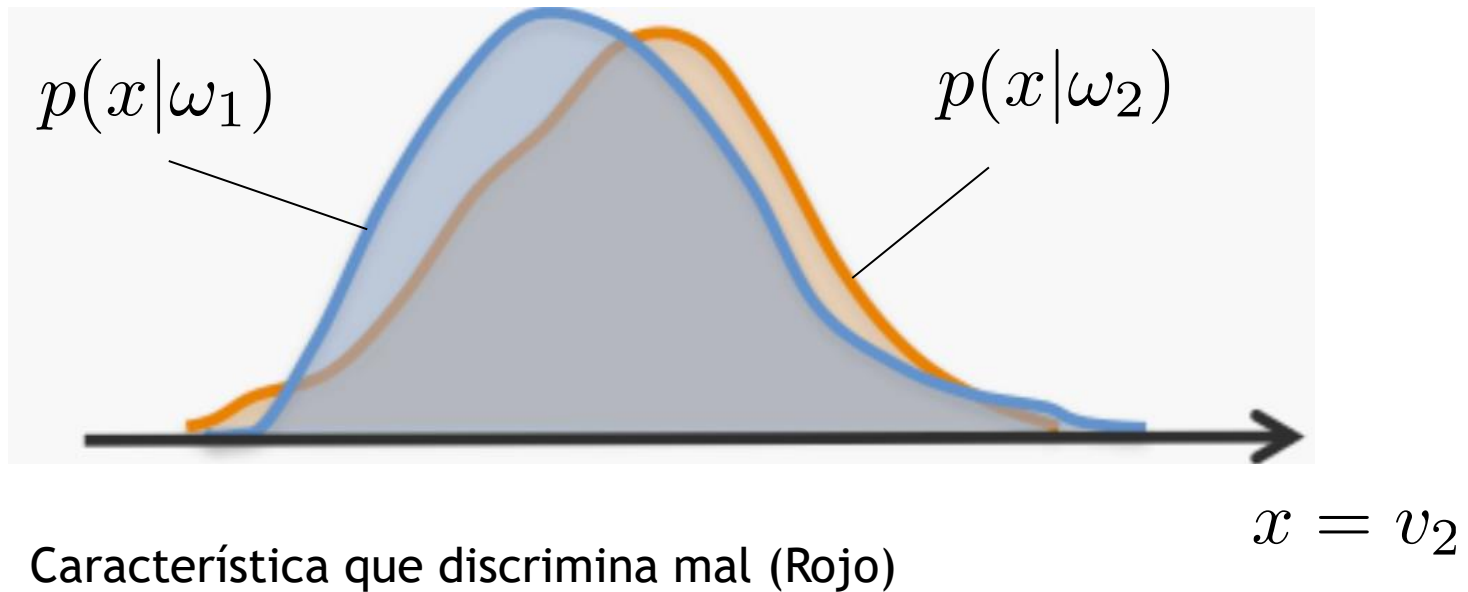
| | | Muestra | A | R | G | B | = V |
|------------|---|----------------------|----------------------|----------------------|----------------------|------|------------|
| Mandarinas | { | 1 | 0.4981 | 0.27 | 0.40 | 0.11 | |
| | | : | 0.3681 | 0.29 | 0.39 | 0.09 | |
| | | 50 | 0.3479 | 0.31 | 0.42 | 0.12 | |
| Naranjas | { | 51 | 0.4601 | 0.35 | 0.38 | 0.13 | |
| | | : | | | | | |
| | | 125 | 0.9751 | 0.30 | 0.73 | 0.10 | |
| | | V₁ | V₂ | V₃ | V₄ | | |

Formulación Metodológica

1. Obtención de la Información Original
2. Pre-procesamiento
3. Extracción de Características
4. Normalización de Características
- 5. Análisis de Características**
6. Selección de Características
7. Diseño del Clasificador
8. Evaluación del Desempeño

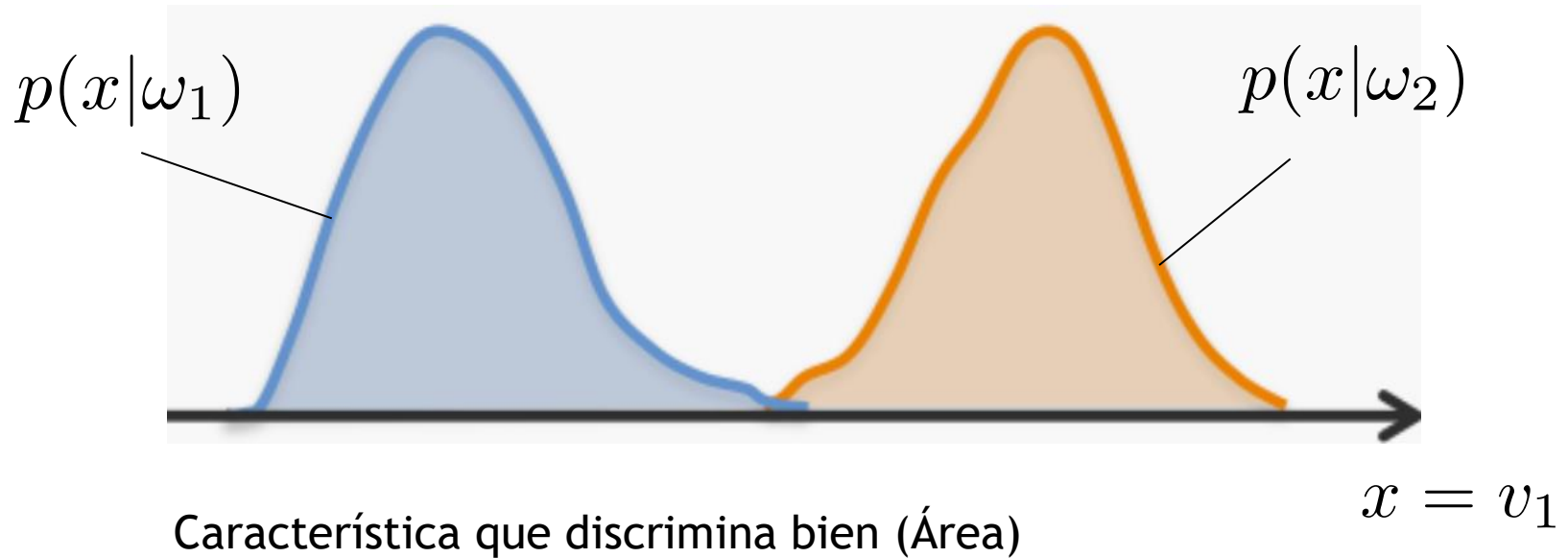
5. Análisis de Características

Histogramas en 1D para dos clases (ω_1, ω_2)



5. Análisis de Características

Histogramas en 1D para dos clases (ω_1, ω_2)

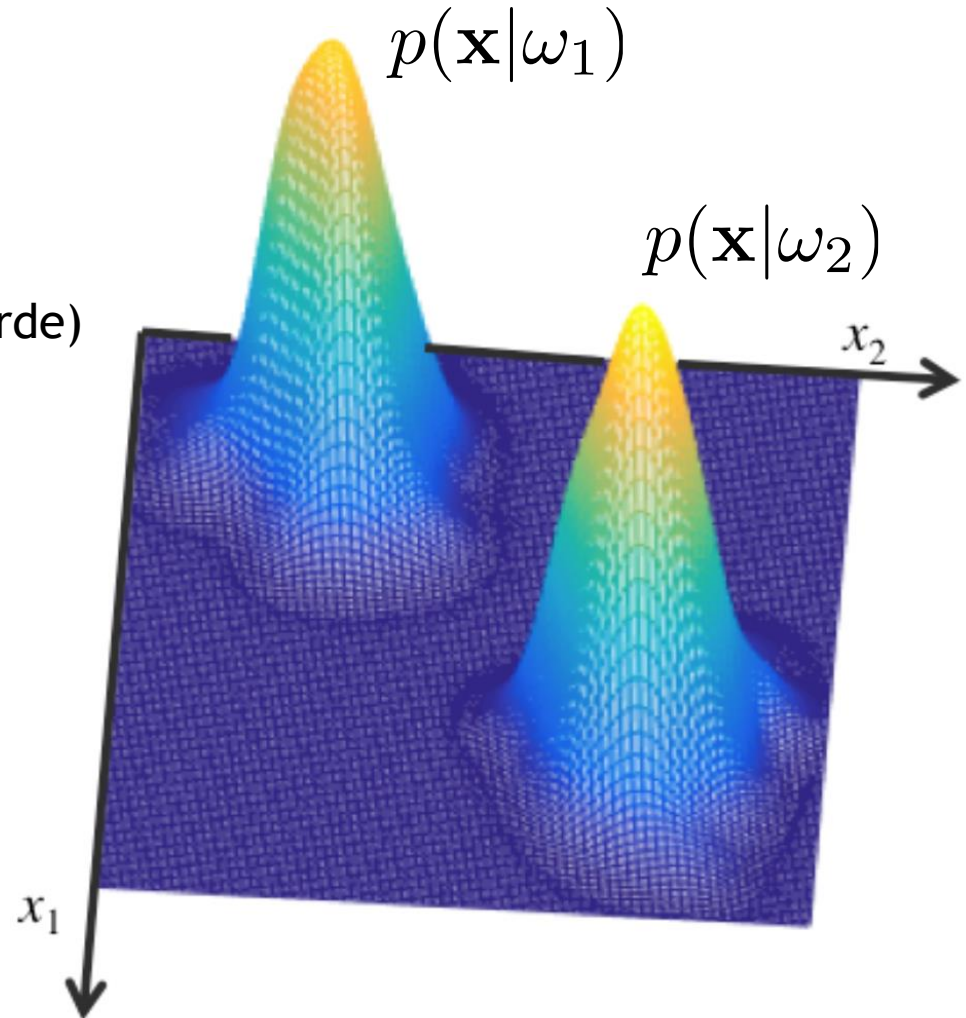


5. Análisis de Características

Histogramas en 2D para dos clases (ω_1, ω_2)

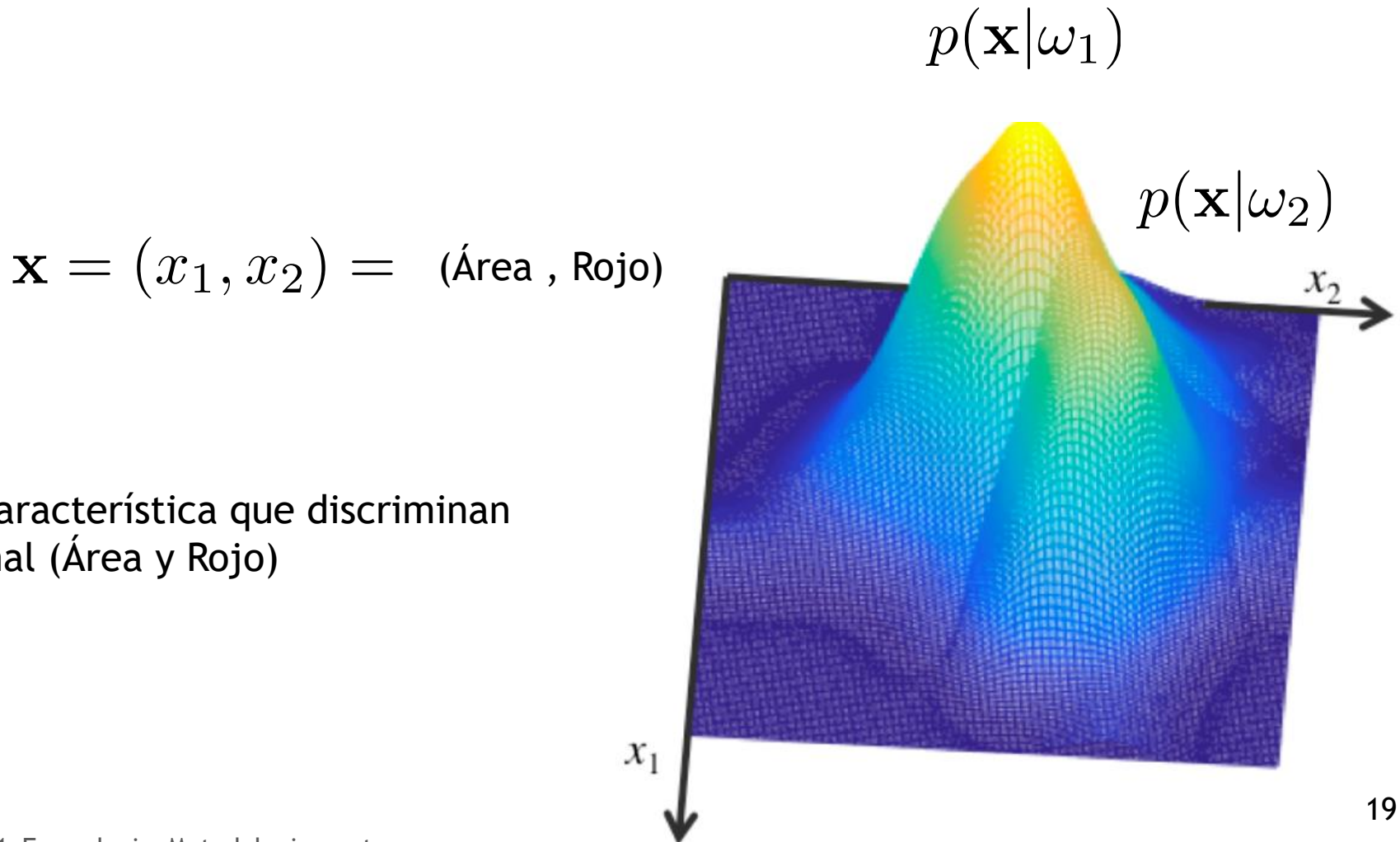
$$\mathbf{x} = (x_1, x_2) = (\text{Área}, \text{Verde})$$

Característica que discriminan bien (Área y Verde)

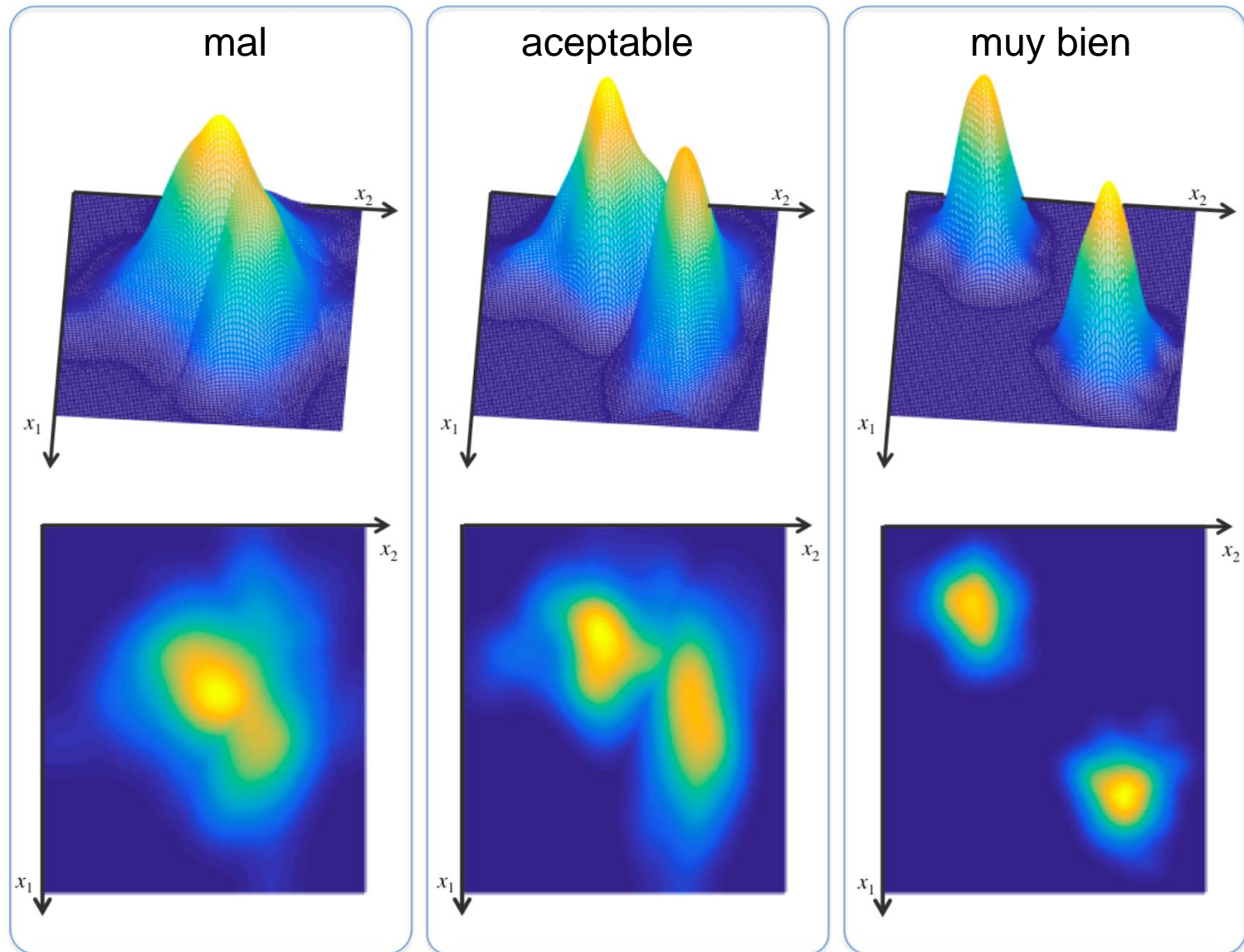


5. Análisis de Características

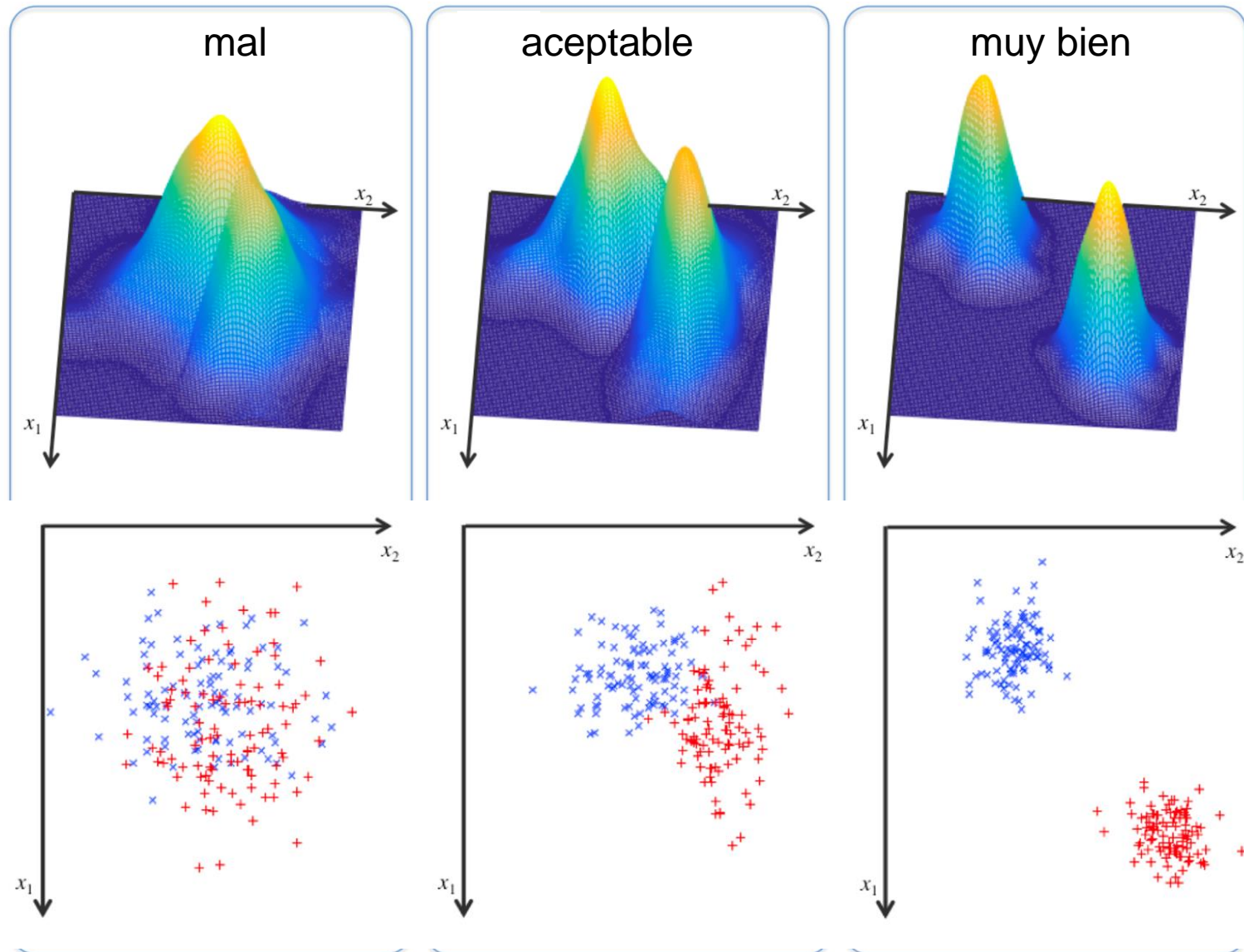
Histogramas en 2D para dos clases (ω_1, ω_2)



5. Análisis de Características



5. Análisis de Características



Formulación Metodológica

1. Obtención de la Información Original
2. Pre-procesamiento
3. Extracción de Características
4. Normalización de Características
5. Análisis de Características
- 6. Selección de Características**
7. Diseño del Clasificador
8. Evaluación del Desempeño

6. Selección de Características

Se escogen aquellas que en conjunto discriminan bien...

| | | Muestra | A | R | G | B | = V |
|------------|---|-----------------------|-----------------------|-----------------------|-----------------------|------|------------|
| Mandarinas | { | 1 | 0.4981 | 0.27 | 0.40 | 0.11 | |
| | | : | 0.3681 | 0.29 | 0.39 | 0.09 | |
| | | 50 | 0.3479 | 0.31 | 0.42 | 0.12 | |
| Naranjas | { | 51 | 0.4601 | 0.35 | 0.38 | 0.13 | |
| | | : | | | | | |
| | | 125 | 0.9751 | 0.30 | 0.73 | 0.10 | |
| | | v ₁ | v ₂ | v ₃ | v ₄ | | |

$\mathbf{Z} \leftarrow$ Columnas escogidas de \mathbf{V}

6. Selección de Características

Se escogen aquellas que en conjunto discriminan bien...

| Muestra | | A | G |
|------------|-----|--------|------|
| Mandarinas | 1 | 0.4981 | 0.40 |
| | : | 0.3681 | 0.39 |
| | 50 | 0.3479 | 0.42 |
| Naranjas | 51 | 0.4601 | 0.38 |
| | : | | |
| | 125 | 0.9751 | 0.73 |

$\mathbf{Z}_1 = \mathbf{V}_1 \quad \mathbf{Z}_2 = \mathbf{V}_3$

$= \mathbf{Z}$

$\mathbf{Z} \leftarrow$ Columnas escogidas de \mathbf{V}

Formulación Metodológica

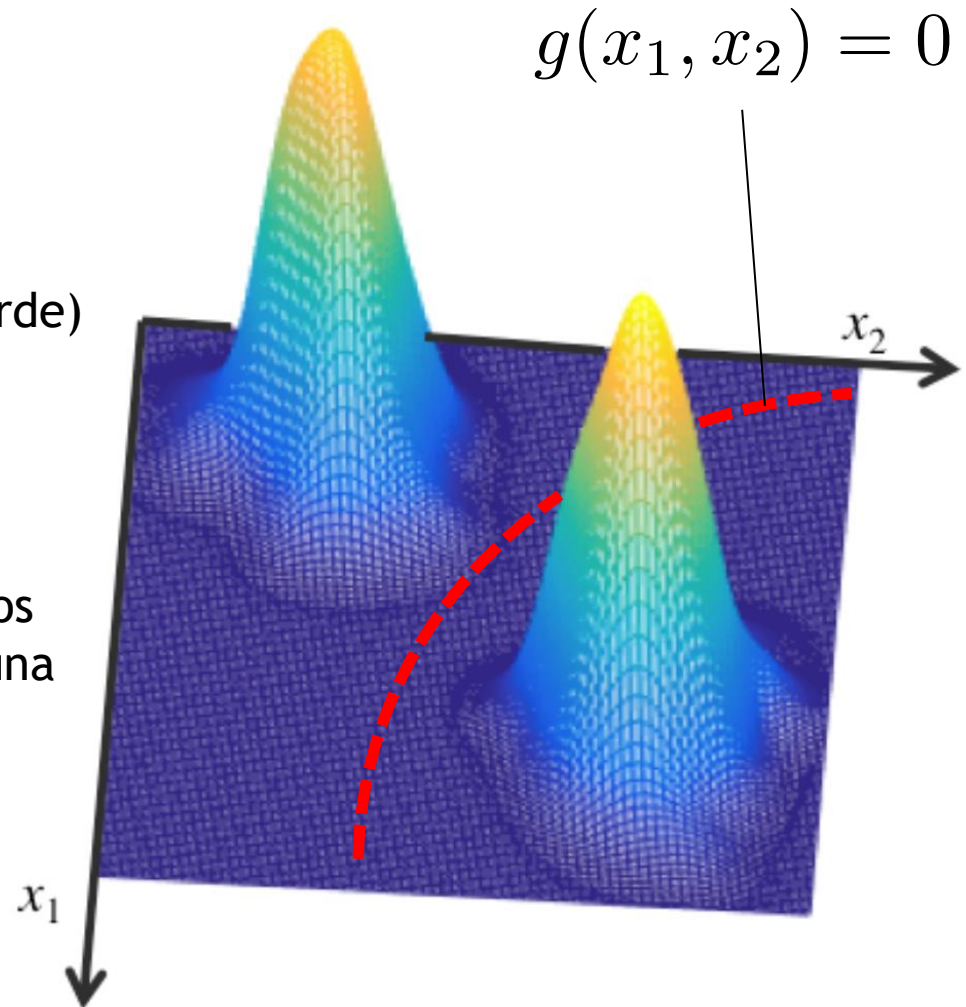
1. Obtención de la Información Original
2. Pre-procesamiento
3. Extracción de Características
4. Normalización de Características
5. Análisis de Características
6. Selección de Características
- 7. Diseño del Clasificador**
8. Evaluación del Desempeño

7. Diseño del Clasificador

Se busca la línea de separación entre las clases

$$\mathbf{X} = (x_1, x_2) = (\text{Área}, \text{Verde})$$

La función g depende de parámetros que deben estimarse minimizando una función de costo



Formulación Metodológica

1. Obtención de la Información Original
2. Pre-procesamiento
3. Extracción de Características
4. Normalización de Características
5. Análisis de Características
6. Selección de Características
7. Diseño del Clasificador
- 8. Evaluación del Desempeño**

8. Evaluación de Desempeño

Se evalúa que tan bien se desempeña el clasificador...

Matriz de Confusión 'C'

| | | Predicción | | | |
|------|------------|------------|------------|-----|------------|
| | | ω_1 | ω_2 | ... | ω_K |
| Real | ω_1 | | | | |
| | ω_2 | | | | |
| | : | | | | |
| | ω_K | | | | |

C_{ij} Número de muestras que pertenecen a la clase 'i'
clasificadas como clase 'j'

8. Evaluación de Desempeño

Se evalúa que tan bien se desempeña el clasificador...

Predicción

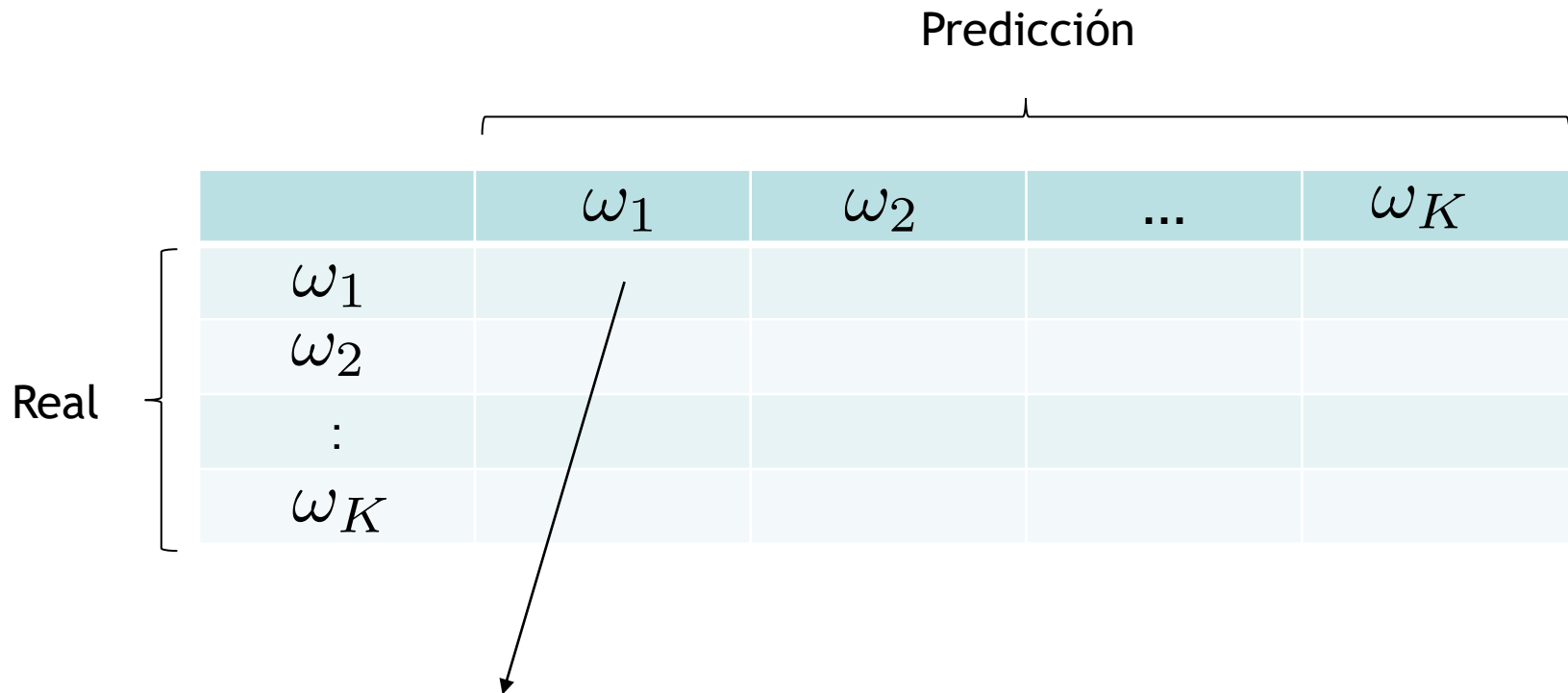
| | | ω_1 | ω_2 | ... | ω_K |
|------|------------|------------|------------|-----|------------|
| Real | ω_1 | | | | |
| | ω_2 | | | | |
| | : | | | | |
| | ω_K | | | | |

Número de muestras que pertenecen a la clase 1
pero que fueron clasificadas como clase 2

8. Evaluación de Desempeño

Se evalúa que tan bien se desempeña el clasificador...

| | | Predicción | | | |
|------|------------|------------|------------|-----|------------|
| | | ω_1 | ω_2 | ... | ω_K |
| Real | ω_1 | | | | |
| | ω_2 | | | | |
| | : | | | | |
| | ω_K | | | | |

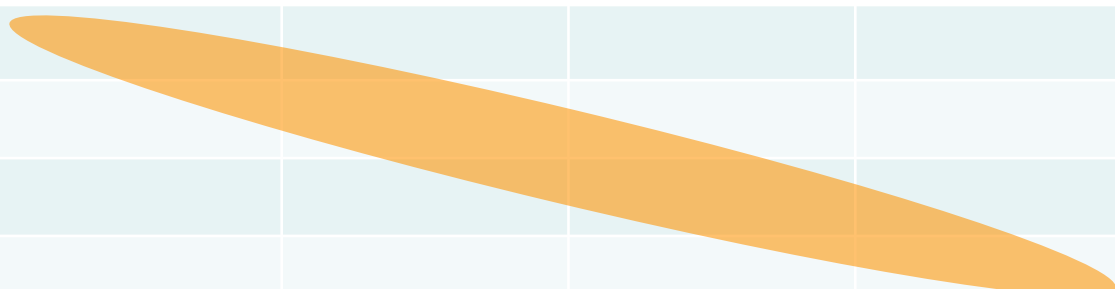


Número de muestras que pertenecen a la clase 1
y que fueron correctamente clasificadas como clase 1

8. Evaluación de Desempeño

Se evalúa que tan bien se desempeña el clasificador...

| | | Predicción | | | |
|------|------------|------------|------------|-----|------------|
| | | ω_1 | ω_2 | ... | ω_K |
| Real | ω_1 | | | | |
| | ω_2 | | | | |
| | : | | | | |
| | ω_K | | | | |



$$\text{Accuracy (Acc)} = \frac{\text{Suma de la Diagonal}}{\text{Total de Muestras}}$$

Existen otras métricas

Formulación Metodológica

1. Obtención de la Información Original
2. Pre-procesamiento
3. Extracción de Características
4. Normalización de Características
5. Análisis de Características
6. Selección de Características
7. Diseño del Clasificador
8. Evaluación del Desempeño